Expanding Explainability: Towards Social Transparency in AI systems

Abstract

As AI-powered systems increasingly mediate consequential decision-making, their explainability is critical for end-users to take informed and accountable actions. Explanations in human-human interactions are socially-situated. AI systems are often socio-organizationally embedded. However, Explainable AI (XAI) approaches have been predominantly algorithm-centered. We take a developmental step towards socially-situated XAI by introducing and exploring Social Transparency (ST), a sociotechnically informed perspective that incorporates the socio-organizational context into explaining AI-mediated decision-making. To explore ST conceptually, we conducted interviews with 29 AI users and practitioners grounded in a speculative design scenario. We suggested constitutive design elements of ST and developed a conceptual framework to unpack ST's effect and implications at the technical, decision-making, and organizational level. The framework showcases how ST can potentially calibrate trust in AI, improve decision-making, facilitate organizational collective actions, and cultivate holistic explainability. Our work contributes to the discourse of Human-Centered XAI by expanding the design space of XAI.